Author Names & Affiliations

- Neil Heffernan - Worcester Polytechnic Institute
- Anthony Botelho - Worcester Polytechnic Institute

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Education, Technology in Education, Computer-Based Learning Platforms

Title of Submission

Research Around How to Design Crowdsourcing Systems

Abstract (maximum ~200 words).

An increasing number of educators are incorporating computer-based systems into the classroom to augment existing instructional practices. Students interact with these systems to complete homework and classwork, while the teacher is provided with reports of student progress that lead to data-driven decisions regarding the most impactful topics to focus on in class as well as what remedial practice is need to benefit the students. While countless teachers devote copious amounts of time building and designing content for their students, no platform exists to share this content with other teachers who may find their students struggling with similar topics. Similarly, existing problems within the textbook used by teachers lacks the feedback and student aid that is often provided from content developed within learning systems. What is strongly needed within the field of education is funded research on how to create products that leverage crowdsourcing. What are the models and modes that allow success to be had? This requires interdisciplinary work between computer scientists, software engineers, designers, and machine learning experts to build systems that enable crowdsourcing to help create better content to benefit students.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

Crowdsourcing is emerging as a powerful way for products to develop. Wikipedia is a simple example, but another, Captcha, using humans to transcribe text, is perhaps a less appreciated example. Stackoverflow, yet another example, crowdsources problems and the answers to those problems with remarkable effectiveness. In terms of the education domain teacherpayteachers.com are BetterLessons.com are two examples of companies trying to crowdsource from teachers content. But crowdsourcing is in its infancy and for it to be deployed well, a lot of research needs to be done.

While many teachers have adopted the use of computer-based systems as the medium for students to complete homework and classwork, the benefit of using these systems is limited by the content available. It is often the case that these computer-based systems, often also referred to as intelligent tutoring systems (ITS) as they often employ methods of providing aid and feedback for students as they work, contain a wealth of problems on various topics on a field of focus. In math domains, the problem begins to become more apparent. Systems like Cognitive Tutor Reasoning Mind, Agile Mind and ASSISTments host a large amount of expert-generated content, but teachers often find themselves wanting to assign more of a particular subject matter that may be limitedly available within a given system. This is not to say that there is a lack of content within these systems, but there exists a bottleneck in that while domain experts are generating problems covering a breadth of topics, teachers realize the specific content that their students are struggling with based on reports of progress within these systems, and it is from these teachers that more, directed content can be generated and shared.

Consider the scenario in which a teacher gives a homework assignment and finds that his/her students are struggling on, for sake of example, solving two-step equations. The system he/she uses in the classroom has a handful of problems pertaining to this subject, but have all just been used on the assignment. The teacher in this scenario may find more problems of this type in the students' textbooks, or may even create his/her own content. Learning platforms, like ASSISTments, provide the means for teachers to create their own content within the system, and also make it easier for teachers to assign existing problems from many supported textbooks, although these lack the tutoring aid available in other hosted content.

The described scenario exemplifies several intersecting needs of the teacher. First, the need for more "similar" content. Teachers often want to find content comprised of similar material, with enough variation to gauge student understanding of the intended topic. Second, there is a need to share content that has been found or created by the teacher. If a teacher finds that a particular set of problems was largely beneficial to his/her students, it will likely be useful for other teachers as well. Being able to identify the best content, as decided by teachers who assign such problems, is further important in helping teachers focus in on the material that will best help students learn. (The conference called HCOMP (www.HumanComputation.com) has a community of researchers in design and machine learning working in this areas) Third, and perhaps most importantly, in addition to being able to share content created or found by the teacher, the ability to build upon existing problems from the already-available textbook is at the foundation of the previous described needs of the teacher.

Particularly in math education, teachers utilize a common set of textbooks. These textbooks often form an integral part of the curriculum and supply numerous problems, often grouped by content and type that are assigned to students year after year. Even with the adoption of computer-based systems within the classroom, textbooks are still an important source of supplementary instruction and coursework. This disconnect between textbooks and ITS is problematic as each is likely to benefit the other in regard to improving student learning; textbooks provide a large amount of organized content, while the ITS provide additional student aid and provide detailed reports to teachers on student progress.

It is for this reason, remedying the disconnect that exists between content and available aid and feedback, in addition to the inability to share such content amongst teachers, that research is needed into how to best collect and share content, in a crowd-sourced manner, from teachers to benefit student learning. A framework is needed to support and promote the best content available to help students. This teacher-sourced content, seeded by the organized content that already exists in textbooks, needs to become available to other teachers and iteratively improved upon with other teacher-sourced feedback and the creation of similar problems to be grouped together.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

There are a number of considerations that must be made in pursuing the posed research challenges in regard to the infrastructure required to support teacher-sourced content. This challenge proposes a need to build workflows and frameworks around the idea of allowing teachers to build and share new or variations of existing content, student aid and feedback, and iteratively improve on existing content with that aid and feedback. Such a framework would allow teachers to easily find similar content to assign to students in need of remedial practice, and also provide an easy means of providing the best content as determined by teachers who have previously assigned it.

The posed research challenge is made with existing tutoring systems in mind. Previously mentioned, ASSISTments is one such learning platform that hosts a wealth of problems, both expert-generated and teacher-generated, but lacks the means of allowing teachers to easily

share this content.

What is proposed as the needed infrastructure is exemplified by Wikipedia. All information there is crowd-sourced by its users. Users are able to create new content, add to existing content, and even edit and alter information on each page of the website. In the paradigm of education, a similar workflow is needed to host content as well as feedback for that content. For example, a teacher may find a small set of problems from his/her textbook that is found to be difficult for students. Realizing this, a number of similar problems is created for the purpose of remedial practice. These new problems could be then be shared on the proposed platform for other teachers to find. Another teacher may find that teacher-created content and realize that he/she could improve upon it by including feedback messages when students are in need of aid, sharing that feedback as well. A third teacher may then find the content and change some of the feedback based on a different instructional approach. In the described scenario, three versions of the content exist, and could presumably even be accompanied by summaries of how beneficial teachers and students found the material. In this way, the content is iteratively improved, in a data-driven manner, allowing teachers to select what is found to be the best content. As a teacher-sourced system, it would be the hope that the amount of content available would scale with each teacher and also with each textbook presumably forming the foundation on which content can be organized and easily found.

While the infrastructure needed to host and version the content as described above forms one challenge, limiting accessibility to teachers is another needed consideration. The posed research challenge is beyond simply developing open-source content to be shared, as it is most beneficial to the students for that content to be shared strictly between teachers and administrators. Eureka Math, formerly known as EngageNY, allowed free-use access to a curriculum, comprised of a number of problems and content that had found to be beneficial in helping students learn in the state of New York. This content has been widely-adopted in classrooms and even into computer-based systems such as ASSISTments. As an open-source, widely used curriculum, however, students can all-too-easily find answer keys to the assigned problems; this can detract from their own process of learning by exploiting this availability of answers. For this reason, further infrastructure is needed to host and share teacher-sourced content while similarly restricting access to teachers and administrators, deterring students from exploiting this freely-available material.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

As described above, the current posed challenge addresses the need of teachers to be able to find, share, and iteratively improve on content for the benefit of their students. Optimistically, such an infrastructure could lead to an abundance of teacher-sourced content. A further challenge exists in this particular case in how to help teachers find and identify the best content. For reasons of scalability, the posed challenge has been described as user-managed, similar to Wikipedia, however it may be helpful to understand the best design that merges content with data representing student performance, teacher opinion, or other such statistics such that teachers who find large quantities of content can make data-driven decisions regarding what to assign.

Thus far, the problem has also been posed within the context of specific learning platforms. It may be beneficial to teachers for developers to further consider the generality of such a sharing platform. Specifically, if teachers find content outside of a particular learning platform, allowing teachers to easily port that content to share with others is itself important. While this could refer to other systems, specifically this consideration is made in regard to textbooks. As new textbooks are implemented in classrooms, or if problems or similar problems to those found in a textbook do not exist on the sharing platform, it should be easy for teachers to include this information. This particularly becomes challenging to those teachers who do not commonly use technology in such a manner. Understanding the usability of such workflows and developing such a platform for a variety of users from various backgrounds is a further consideration needed toward the posed research.

A final consideration is that of content ownership. It is likely beneficial to students to be able to share the best content freely with other teachers. However, it is further understood, particularly in the cases of textbook content, that considerations must be made as to not infringe upon existing copyrights. Permissions to use and share content must be gained before being able to do so, and should be considered as part of the workflow to build and share such content. Many, if not all, computer-based systems already consider this aspect of permissions, and would simply need to be a consideration in a teacher-sourced platform as well.

Workshops by a small group of researchers interest in doing research on how to crowdsourcing have been well attended at various venues (such as 1) CHI'16 Human Factors in Computing Systems., 2) Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing, & 3) the Second AAAI Conference on Human Computation and Crowdsourcing.) There

is a growing groups of researchers interested in this phenomenon, but NSF should focus attention on the unique benefits of crowdsourcing in education.

Consent Statement